

Implementation Support Improves Outcomes of a Fluency-Based Mathematics Strategy: A Cluster-Randomized Controlled Trial

Owen, Kaydee; Hunter, Stacey; Watkins, Richard; Payne, Joshua; Bailey, Tom; Gray, Cameron; Hastings, Richard; Hughes, Carl

Journal of Research on Educational Effectiveness

DOI:

[10.1080/19345747.2021.1875526](https://doi.org/10.1080/19345747.2021.1875526)

Published: 03/07/2021

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Owen, K., Hunter, S., Watkins, R., Payne, J., Bailey, T., Gray, C., Hastings, R., & Hughes, C. (2021). Implementation Support Improves Outcomes of a Fluency-Based Mathematics Strategy: A Cluster-Randomized Controlled Trial. *Journal of Research on Educational Effectiveness*, 14(3), 523-542. <https://doi.org/10.1080/19345747.2021.1875526>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Implementation Support Improves Outcomes of a Fluency-Based Mathematics Strategy: A Cluster-Randomized Controlled Trial

Kaydee. L. Owen^{a*}, Stacey. H. Hunter^a, Richard. C. Watkins^{a,b}, Joshua. S. Payne^a, Tom Bailey^c, Cameron Gray^d, Richard. P. Hastings^c, and J. Carl Hughes^a

^a Collaborative Institute for Education Research, Evidence, and Impact (CIEREI), Bangor University, Wales, United Kingdom; ^b Regional School Effectiveness and Improvement Service for North Wales (GwE), Wales, United Kingdom; ^c Centre for Education Development, Appraisal and Research (CEDAR), University of Warwick, England, United Kingdom; ^d School of Computer Science and Electrical Engineering, Bangor University, Wales, United Kingdom

*Correspondence email: kaydee.owen@bangor.ac.uk

Implementation Support Improves Outcomes of a Fluency-Based Mathematics Strategy: A Cluster-Randomized Controlled Trial

The Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS) strategy promotes fast and accurate recall. The existing literature suggests that the strategy can help learners improve academic outcomes. Through a cluster randomized controlled trial, we assessed the impact of implementation support on children's mathematics outcomes during a teacher led SAFMEDS intervention. Following training and prior to baseline assessments, we randomly allocated schools to receive either no ($n = 31$) or ongoing ($n = 33$) support from a researcher. Support consisted of three in-situ visits and email contact. Assessors remained blind to the condition of the schools throughout. We analyzed the outcomes of children ($n_{\text{Support}} = 294$, $n_{\text{NoSupport}} = 281$) using a multi-level mixed effects model; accounting for the children nested within schools. The results suggest that implementation support has a small effect on children's fluency of arithmetic facts (Mathematics Fluency and Calculation Tests (MFaCTs): Grades 1-2, $d = 0.23$, 95% CI: 0.06 to 0.40; MFaCTs: Grades 3-5, $d = 0.25$, 95% CI: 0.08 to 0.42). These results are larger than the average effect sizes reported within professional development literature that apply coaching elements to mathematics programs.

Keywords: fluency; mathematics; coaching; randomized controlled trial

Introduction

Evidence-based practice has the potential to generate effective educational programs and promote positive outcomes for students. Yet, Klingner, Boardman, and McMaster (2013) outlined that educators are less likely to adopt and sustain these practices at scale without strategic and systematic support from researchers. Implementation science suggests that researchers first conduct efficacy studies on a small scale to validate an intervention before implementing them at a larger scale under real-world (day-to-day) conditions. The quality of implementation during the latter phase may bound the benefits of evidence-based programs in school environments (Cook & Odom, 2013). Implementation fidelity refers to the extent to which someone implements a program according to the original and intended design (Lee, Penfield, & Maerten-Rivera, 2009). Durlak and DuPre (2008) found that educators who do not specialize in research (e.g., teachers) often do not implement an intervention to 100% fidelity under the real-world conditions of a classroom. They also found that low-quality implementation of evidence-based interventions results in smaller effect sizes on outcome variables; including those linked to student achievement. This highlights the importance of identifying effective implementation support models to ensure that teachers are able to elicit desired and intended outcomes from evidence-based educational interventions.

Training can be an effective way of helping teachers to develop conceptual understanding of interventions but alone may not yield sufficient changes in practice (Education Endowment Foundation, 2019). Coaching teachers offers a lever for improving the quality of implementation by supporting them to translate knowledge into classroom practice (Kraft, Blazar, & Hogan, 2018). Sailors and Shanklin (2010) used the term coaching to describe a process of sustained school-based support from a knowledgeable individual. Coaches model research-driven interventions and work with teachers to explore how they can use the strategies with their own students. Coaching

programs can take a variety of forms, but generally consist of one-on-one interactions between a coach and a teacher. These interactions provide a platform for teachers to receive individualized feedback based on their professional development (PD) needs (Fletcher & Mullen, 2012).

Following a meta-analysis of 60 studies, Kraft et al. (2018) found that teachers often receive coaching in conjunction with additional treatment elements (i.e., in 90% of the reviewed studies teachers received coaching alongside group training, instructional content, and/or video resources). Their analysis revealed a pooled effect size of $+0.18$ standard deviations (*SD*) relating to the effect of these programs on student achievement and $+0.49$ *SD* relating to teachers' instructional practices. In their theory of action, Kraft et al. outlined that training sessions help improve teacher pedagogical and content knowledge. This knowledge, alongside coaching and the availability of relevant materials, positively influences teaching behavior. As a result, teachers implement higher-quality teaching practices and are better able to identify and use strategies that support student outcomes. However, it is worth noting that most of the interventions that met Kraft et al.'s inclusion criteria focused on applying these practices to literacy and content-based interventions; with only two studies reporting the outcomes of mathematics programs. Moreover, Kraft et al.'s analysis revealed that the effects on student outcomes from larger-scale effectiveness trials were smaller ($+0.10$ *SD*) than those employing smaller-scale efficacy designs ($+0.28$ *SD*). Whilst coaching might be a valuable tool, research is still needed to disentangle the effects of coaching from additional treatment elements and to establish the effects of using coaching programs at scale.

In a complementary meta-analysis of 95 studies, Lynch, Hill, Gonzalez, and Pollard (2019) reviewed mathematics and science interventions supported by PD and/or

curriculum materials. Lynch et al. defined PD as a set of experiences that intend to affect change in teacher- and classroom-level phenomena. They too highlighted that PD programs can be, and often are, multifaceted. As such, their inclusion criteria focused on the number of hours a teacher spent experiencing PD; the focus on improving knowledge of content, pedagogy, and/or use of curriculum materials; as well as the format of the program (e.g., one-on-one coaching, summer workshops, online learning). Curriculum materials are instructional practices, guided by activities and text within the program itself. Their review identified that 22% of studies focused on PD alone, whilst 75% used PD in combination with curriculum materials. Overall, they found PD programs to be effective. However, only 20% of the PD programs included a coaching element. There was no evidence that coaching elements added value in terms of outcomes but neither did they reduce intervention's effectiveness. The vast majority of included studies with a coaching element were multi-component programs. With few published studies reporting the outcomes of coaching as a standalone PD format to support mathematics interventions, further research is needed.

SAFMEDS overview and prior research

The current study is set within the context of North Wales, United Kingdom. Following disappointing results in the internationally comparative Programme of International Student Assessment tests in 2009 (OECD, 2010), the Welsh Government identified a need to raise educational standards in their schools. In recent years, education policymakers in Wales have focused on improving the use of evidence-based practice within education (Furlong, 2015; OECD, 2016). As a result, an increasing number of teachers in North Wales are using the Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS) strategy in their classrooms to improve children's fluency of basic mathematics skills (Tyler et al., 2019).

Traditionally, within educational practice teachers deem children to have mastered skills if they are able to perform them to a level of 80-100% accuracy (Fuller & Fienup, 2018). Binder, Haughton, and Bateman (2002) argued that a percentage correct criterion is too simplistic—being accurate is necessary, but not sufficient to demonstrate mastery of content. If children practice skills beyond mastery they will be able to develop fluency (the combination of accuracy *and* speed). Adding a dimension of time into assessment provides more detail about performance and can more accurately predict whether children will be able to retain, apply, and generalize learned skills (Binder, 1996; Johnson & Street, 2012). SAFMEDS is a practice and assessment strategy that applies the principles of precision teaching (PT) to help children develop their skills to fluency (Lindsley, 1995).

Kubina and Yurich (2012) described PT as a system for defining, measuring, recording, and analyzing teaching effectiveness on a child-by-child basis. To achieve this, teachers must reflect upon children's learning regularly, and use these data to make subsequent decisions about their teaching approach (Lindsley, 1995). A child engages with the SAFMEDS strategy using a deck of flashcards, with a question or statement on the front and the corresponding correct answer on the back. They read the front of the card silently before vocalizing the answer (Quigley, Peterson, Frieder, & Peck, 2018). During each 1-minute timing, they aim to get through as many cards as possible, whilst separating their correct responses from their "not yet's" (Cihon, Strutz, & Eshleman, 2012). The child then plots their best score from the session on a Standard Celeration Chart (SCC), which develops a learning picture over time (for more details see Lindsley, 1995). Learning pictures enable teachers, children, and/or other practitioners to decide if additional support is necessary. For example, if a learning picture shows that the number of cards a child can answer correctly in 1-minute has plateaued over

several consecutive sessions, an intuitive approach might be to assess if the child has mastered all of the necessary prerequisite skills, and if not, to ensure that they do so (Johnson & Street, 2012).

The SAFMEDS strategy has clear utility within schools, with a growing quantitative evidence-base suggesting some of the associated advantages. For example, practitioners can adapt the strategy to implement it on a one-on-one basis (e.g., Cunningham, McLaughlin, & Weber, 2012), with small groups of children (e.g., Beverley, Hughes, & Hastings, 2018), or class-wide (Hunter, Beverley, Parkinson, & Hughes, 2016). The strategy also has evidence to support its effectiveness amongst different populations including learners attending mainstream classes and children with additional learning needs (e.g., Casey, McLaughlin, Weber, & Everson, 2003; Greene, Mc Tiernan, & Holloway, 2018; Kubina, Ward, & Mozzoni, 2000).

Much of the available literature on the SAFMEDS strategy documents small *N* and case study research designs. These studies demonstrate the positive effects of the approach in improving academic outcomes of learners across a variety of domains. This includes helping children become more fluent at arithmetic (see for example, Casey et al., 2003), recalling content specific terminology and definition dyads (see for example, Stockwell & Eshleman, 2010), as well as sight reading Dolch words (Lambe, Murphy, & Kelly, 2015). More recently, comparative group studies have investigated the effectiveness of the SAFMEDS strategy against an education as usual control group (e.g., Hunter et al., 2016, Greene et al., 2018). Within these studies, a researcher with experience using the SAFMEDS strategy was present at each intervention session to support implementation and ensure high levels of fidelity.

Although sparse, there is some evidence to suggest that teachers can elicit positive student outcomes from a SAFMEDS intervention even when researchers offer

no in-situ support following training. Beverley, Hughes, and Hastings (2016) acknowledged the importance of providing teachers with the training necessary for them to implement and manage a SAFMEDS intervention. Following training, the teachers participating in their study did not receive any in-situ support from a researcher to implement the strategy on a class-wide scale. The results demonstrated that the class of children who engaged with the SAFMEDS intervention made more reliable fluency progress between pre- and post-test compared to the class of children who did not use the strategy.

To date, the majority of empirical research investigating the SAFMEDS strategy focuses on efficacy designs with researcher driven implementation. Beverley et al's (2016) study suggests that teachers can elicit positive student outcomes under conditions with no researcher input following training. Whilst both of these approaches have shown positive results, it is still unclear whether researcher involvement after initial teacher training is important for implementation and children's outcomes. The aim of the current study was to provide direct insight into the impact of coaching (i.e., in-situ individualized implementation support from a researcher) during a teacher led SAFMEDS mathematics program in schools. The teachers and teaching assistants used the SAFMEDS strategy with the children in their schools to help develop fluency of arithmetic.

In our theory of action, the initial training intended to support teacher's pedagogical knowledge of the SAFMEDS strategy and the associated data-driven teaching practice (PT). We anticipated that coaching support would improve the fidelity of the teacher's implementation. A researcher tailored each in-situ visit to the individual needs of each teacher, but broadly these sessions aimed to address challenges such as: interpreting learning pictures; identifying and correcting children's procedural steps as

they progressed through each SAFMEDS timing; and managing challenges such as cheating and identifying appropriate learning materials. As a result of more accurate implementation, our theory of action proposed that children attending schools where their teacher received coaching would make greater fluency progress between baseline and follow-up, compared to those attending schools that did not receive coaching following training.

In line with previous studies that have investigated the effects of the SAFMEDS strategy, the outcomes from this research relate to children's arithmetic fluency. We acknowledge that it would have been beneficial to collect data directly relating to fidelity of implementation but were unable to due to practical and funding restraints. To our knowledge, this is the first randomized controlled trial (RCT) investigating the effects of providing implementation support to teachers following SAFMEDS training. Answering this question would provide a foundation for further research investigating the mechanisms that make SAFMEDS coaching programs effective and contribute to the broader literature about the effects of coaching for teachers on intervention outcomes for students.

Method

Trial design and participants

As part of a wider initiative to improve numeracy standards across North Wales, the Regional School Effectiveness and Improvement Service for North Wales (GwE) disseminated the initial advertisement for this project. For a school to be considered eligible, they had to be located within one of the six local authorities supported by GwE (Conwy, Denbighshire, Flintshire, Gwynedd, Anglesey, or Wrexham). Table 1 outlines the characteristics of the schools included in the randomization.

To participate schools needed to be willing to release teacher(s) to attend the training at the beginning of the project. The nominated teacher needed to be able to invest the necessary amount of time per week to deliver the SAFMEDS intervention (i.e., three 20-minute sessions). The advertisement explained that by enrolling on the project schools would be randomized to one of the two trial arms. Schools had nominated teachers to complete the training *before* they knew which trial arm they had been allocated to. Therefore, any trial arm differences in the roles of teaching staff selected for training by the schools were due to chance. Table 2 displays the baseline characteristics of the teachers who attended training.

Each school selected up to 10 children to participate in the SAFMEDS mathematics intervention prior to randomization. We disseminated an opt-out consent form to all the children's parents/guardians detailing the aims of the study. This form asked if we could collect and analyze their child's outcome data. In instances where the consent form was returned, teachers could still include the children in the SAFMEDS intervention, but we did not collect their anonymized data for analysis. We had consent to analyze the data from 575 children ($n_{\text{Support}} = 294$, $n_{\text{NoSupport}} = 281$), across 60 schools ($n_{\text{Support}} = 31$, $n_{\text{NoSupport}} = 29$).

For children in Year 3 or above (aged ≥ 7 years), we asked teachers to implement the intervention with children who scored less than 100 standard points on the national numeracy procedural test undertaken at the end of the preceding academic year. All children in Years 2-9 (aged 6 to 14 years) who attend a maintained school in Wales (i.e., schools funded by a local education authority) sit this formative test at the end of each academic year. Children sit the procedural test online as it offers a personalized assessment experience (i.e., the questions get easier/more challenging

depending on the child's ability). The procedural numeracy test assesses all relevant aspects of the numeracy curriculum in Wales.

Children in Year 2 had not completed the national tests at the start of the study. In these instances, we asked schools to identify the children who they felt needed intervention support to improve fluency of basic mathematics skills and/or who they judged to be working below the expected standard for their age. These children were those who needed supplementary tuition to improve their fluency of arithmetic facts.

The mean age of the children attending schools randomized to the no support arm was 7-years 3-months (range: 6-years 0-months to 9-years 2-months; $SD = 14.34$ months). The mean age of children attending schools allocated to the ongoing support arm was also 7-years 3-months (range: 6-years 0-months to 15-years 10-months; $SD = 14.32$ months). It is worth noting that two secondary (high) schools participated in this study. One of these schools worked with a group of Year 7 students (aged 11-12 years) who significantly underperformed on the procedural test. The other secondary school was a special educational needs school that supported children aged 11-17 years; these children lacked basic mathematics skills (e.g., single digit addition). Table 3 outlines the characteristics of the children included in the randomization. Table 4 displays the baseline characteristics for the children's outcome measures.

Randomization

Randomization occurred after all teachers received the SAFMEDS training but prior to the children completing the baseline assessments. A statistician—who was independent to the study—randomly allocated schools to one of the two trial arms using minimization. During this allocation, the statistician stratified schools by County (local education authority) and the language used predominantly for teaching (English versus Welsh medium). Some of the schools had the same headteacher; in these instances, the

statistician treated the schools as one cluster to prevent bleeding effects across conditions. In terms of hierarchical structure, teachers and children were nested within each school. The first author could not be masked to the randomization due to the need to conduct support visits. However, the assessors who conducted the baseline and follow-up assessments remained blind to the allocation of each school.

[Insert Table 1 here]

[Insert Table 2 here]

[Insert Table 3 here]

[Insert Table 4 here]

Intervention***SAFMEDS training (all teachers)***

All teachers received the same training prior to randomization. During the 3-hour training session, we introduced the teachers to some of the basic theory behind the SAFMEDS strategy, modelled the procedure (as detailed in Table 5), and gave them the opportunity to practice using the cards. Following four 1-minute SAFMEDS timings, we showed the teachers how to record data and graph it on a SCC. During the training, we also emphasized the importance of interpreting learning pictures in relation to children's learning progression throughout the intervention (see Lindsley, 1995). After showing the different learning pictures, we went through a series of common scenarios using SCC data from previous research projects. The scenarios prompted discussion relating to cheating, identifying skill deficits, and deciding whether something in the surrounding environment may be affecting a child's scores (e.g., missing their favorite lesson to take part in the SAFMEDS session, or a loud music lesson scheduled in the room next door). We discussed what learning pictures may develop as a result of these scenarios and suggested some interventions that might be appropriate to try (e.g., creating individualized score targets, building fluency of prerequisite skills, or changing the time/location of the SAFMEDS session).

Throughout the intervention period the children engaged with the SAFMEDS strategy via a deck of flashcards. On the front of each card was a question (e.g., $5 + 6 =$) and on the back was the corresponding correct answer. During the training, we provided teachers with all the materials that they would need to start the SAFMEDS intervention in their school. This included decks of addition and subtraction SAFMEDS cards, score tables, SCCs, and a placemat (so the children could easily distinguish between their "correct" and "not yet" cards). All teachers who attended the training also had access to

printable PDF materials of component arithmetic skills across the national curriculum which they could download at their convenience. We instructed all schools to focus on single digit addition skills first and then progress through card decks as required (in line with the children's learning pictures).

We instructed teachers to use the SAFMEDS strategy at least three times per week with the children they were supporting. Each session should consist of four SAFMEDS timings and last approximately 20-minutes. Within these sessions, the teachers and children had clearly defined roles. The children were to work through their cards independently during each 1-minute timing (as outlined within Table 5). Teachers were required to monitor aspects of fidelity (e.g., ensuring the children: followed each of the appropriate steps, were not cheating, and were regularly engaged with the sessions). Additionally, we encouraged teachers to support children during the error correction step (including some one-on-one or small group teaching if necessary), review charted data regularly, and ensure that children were practicing a skill that was appropriately matched to their existing skill level. Once children had claimed they had become fluent at a deck of cards, it was also important that the teacher was able to verify this (e.g., watch a timing) before they issued a deck for a more difficult skill.

[Insert Table 5 here]

Ongoing support

Schools allocated to the ongoing support trial arm received three in-situ support visits from the first author throughout the duration of the study (November, February, and May). The first author had several years' experience of using the SAFMEDS strategy in schools, so was able to advise teachers on themes around implementation and interpreting the children's data. Each visit was individualized based on the needs of each teacher and the children they were supporting. Examples of support varied, but largely consisted of the following: modelling sessions; observing the teachers delivering the intervention and providing direct feedback on implementation; suggesting interventions for children who were struggling to progress with particular decks (e.g., focusing on building fluency in prerequisite skills); discussing ways that teachers might be able to integrate the intervention more readily (e.g., adopting a peer-led approach to support error correction and reduce cheating); and supporting teachers to interpret the children's learning pictures. Each scheduled visit lasted 1-hour.

Between visits, teachers could email the first author about any issues relating to the intervention or the technology used to support the project. Teachers allocated to this trial arm contacted the author about accessing materials (18 instances; 12 schools), to gain advice about helping children progress (7 instances; 6 schools), and for advice about interpreting data (2 instances; 2 schools).

We gave all teachers—irrespective of trial arm— the option to plot the children's data using either paper or electronic SCCs. There were 10 instances (across 8 schools) where teachers allocated to receive support emailed the first author to report issues logging the data electronically. Moreover, we made all of the resources for this project available via the Welsh Government's online school platform for educational

resources (HwB). Two teachers (across 2 schools) allocated to the support trial arm emailed about gaining access to the SAFMEDS HwB platform.

No support

Following training at the beginning of the project, schools allocated to the no support trial arm received no implementation support from the first author. Teachers in this condition could contact the first author if they had any technical problems accessing the resources or inputting data into the electronic charts, but were not able to seek advice regarding the day-to-day implementation of the SAFMEDS strategy. There were 9 instances (across 7 schools) where teachers emailed the author to request access to resources, 2 instances (across 2 schools) where teachers needed support accessing the SAFMEDS HwB platform, and 10 instances (across 9 schools) where teachers reported issues logging their children's data electronically.

During the training, we highlighted an additional caveat about the support we could offer schools allocated to this arm. We had an ethical obligation to provide the teachers with support if they felt like they could not initiate or sustain the intervention without it. No school in this condition asked for additional support, but if they did, we would have provided it and handled their data appropriately. It is also important to note that the “no support” group was essentially a “support as usual” group in the context of school improvement efforts in Wales. Typically, schools would seek a training course, send their staff on the course, and then implement interventions on their own (unless they specifically purchase additional support with implementation). Thus, we believe that the no support trial arm is an ecologically valid comparison for inclusion within this study.

Procedure

Baseline assessments

The children completed the Mathematics Fluency and Calculation Tests (MFaCTs; Reynolds, Voress, & Kamphaus, 2015). The Grades 1-2 fluency assessment measures addition and subtraction fluency and is intended for children aged between 6-years 0-months and 8-years 11-months. The Grade 3-5 fluency assessment measures addition, subtraction, multiplication, and division fluency; this assessment is intended for children aged between 8-years 0-months and 11-years 11-months. We used both measures with all the children in the sample to provide an inclusive overview of their skill progress across the intervention. To reduce practice effects, the MFaCTs assessments offer parallel test forms. The published statistics for these tests show high internal reliability across ages ($\alpha > .80$). We used Form A during the baseline assessments.

The children came out in a group to complete these assessments but filled in their forms individually and in silence. The assessors provided each child with a pencil and the test form. The children completed MFaCTs: Grades 1-2 first. They had 5-minutes to answer as many of the 100 questions on the page as they could; working across the page from left to right. If they did not know the answer to a question, they were allowed to skip it and move onto the next one. Once the timer finished, an assessor instructed the children to turn the form over so that they could collect them. The children then repeated this procedure for the MFaCTs: Grades 3-5 assessment.

Eight-month follow-up assessments

Eight months post-randomization, we reassessed the children who participated in the study. This process mirrored the administration of the baseline assessments, with the children completing both MFaCTs fluency assessments (Form B).

Figure 1 outlines the flow of participants from enrolment to the final analysis. Prior to the follow-up assessments, four schools indicated that they were no longer using the SAFMEDS intervention due to unforeseen challenges with staffing. Three of these schools were happy for us to still collect follow-up data from their children (denoted as intend to treat); whilst one school was unable to accommodate this (denoted as withdrawal).

[Insert Figure 1 here]

Analysis

The data for this study falls within two hierarchical levels (level 1 = children, level 2 = school). Due to children being nested within schools, we analyzed the data using a multi-linear mixed effects model. This analysis is consistent with other studies that have adopted cluster RCT designs (see, for example, Jahoda et al., 2017; Zimmermann et al., 2014). Linear mixed-effect models enable analysis of continuous outcome variables within hierarchical research designs by partitioning the overall variance of the outcome variable into factors that correspond to the different levels of the hierarchy (Galecki & Burzykowski, 2013). Baayen, Davidson, and Bates (2008) further outlined some of the advantages of using mixed effects modelling over univariate alternatives, such as ANOVA or ordinary least squares regression.

Due to lack of availability of standardized scores for the range of ages included within this sample, we opted to analyze the children's raw scores on the MFaCTs measures. We used Stata v13.0 to analyze the raw data from this trial. Using Xtmixed, we assessed the interaction between time (baseline versus follow-up) and trial arm (ongoing support versus no support) across the fluency (MFaCTs) measures. Level 1 within our model contains covariates associated with individual children, these were: gender, predominant home language, eligibility for free school meals status (eFSM), and school year group. Level 2 within our model refers to the covariates associated with each school, these were: school administrative county, trial arm, and time. The model also generated the intraclass correlation coefficients values (ICCs) associated with each level of the model.

To assess the impact of support, we calculated a Cohen's *d* effect size for each measure. To calculate Cohen's *d* and the associated 95% confidence intervals we adhered to Feingold's (2015) formulae. We have discussed the outcomes of the results

in relation to Cohen's (1988) benchmarks, whereby an effect is small ($d = 0.20$ to $d = 0.49$), medium ($d = 0.50$ to $d = 0.79$), or large ($d \geq 0.80$).

We carried out sensitivity analyses by repeating the main analysis using multiple imputation and a complete cases analysis approach. The effect sizes varied minimally (refer to supplementary material). Existing published data suggest that certain factors predict an attainment gap between sub-groups of school-aged children. These include differences in outcome variables across genders and levels of social deprivation (OECD, 2012). Moreover, Van Rinsveld, Dricot, Guillaume, Rossion, and Schiltz (2017) provided evidence to suggest that bilingual individuals rely on differential activation patterns in the brain to solve simple and complex arithmetic questions in their different languages. As such, we also conducted a series of moderation analyses to investigate the effects of these variables (refer to supplementary material). We found no evidence of these factors moderating the effect of trial arm on children's mathematics outcomes.

Results

Support model

Figure 1 outlines the number of schools who completed each support visit. By the final visit, two schools allocated to receive ongoing support had stopped using the SAFMEDS strategy due to unforeseen changes to staff availability. Of the schools continuing to use the SAFMEDS strategy, all but one engaged with the three support visits. Seventeen schools allocated to the support arm made email contact with the first author between visits to access further support.

Fluency outcomes

We were interested in investigating whether implementation support from a researcher could help improve children's fluency outcomes during a teacher led

SAFMEDS intervention. In terms of the MFaCTs: Grades 1-2 assessment, the statistical analysis revealed a small positive effect of ongoing support over no support on the children's addition and subtraction fluency between baseline and follow-up (Trial arm x Time: $\beta = 2.92$, $SE = 0.86$, $p = .001$, $d = 0.23$). A pairwise comparison of marginal linear predictions, with Bonferroni correction, revealed significant improvements on this measure for children in both arms. Children's raw scores in the support arm improved to a greater extent on average between baseline ($M = 12.00$) and follow-up ($M = 22.59$; $p < .001$) compared to children in the no support arm ($M_{\text{baseline}} = 9.02$, $M_{\text{follow-up}} = 16.50$, $p < .001$).

Analysis of the MFaCTs: Grades 3-5 showed that ongoing support has a small positive effect, relative to no support, on the children's addition, subtraction, multiplication, and division fluency ($\beta = 2.68$, $SE = 0.75$, $p < .001$, $d = 0.25$). Bonferroni-corrected pairwise comparisons revealed that children's raw scores on this measure improved significantly in both arm of the study. Children in the ongoing support arm improved to a greater extent between baseline and follow-up ($M_{\text{baseline}} = 8.52$, $M_{\text{follow-up}} = 19.12$, $p < .001$) than children in the no support arm ($M_{\text{baseline}} = 5.94$, $M_{\text{follow-up}} = 13.76$, $p < .001$). Table 6 displays further descriptive statistics from the linear mixed effects analysis for both MFaCTs outcomes.

[Insert Table 6 here]

Discussion

Our aim for the current study was to gain insight into the putative benefits of providing teachers with implementation support throughout a SAFMEDS mathematics program. An increasing number of teachers across North Wales are using the SAFMEDS strategy to support children's fluency of basic mathematics skills. Yet, no known research internationally had investigated whether implementation support from a researcher can lead to better fluency outcomes than the more traditional "no support" approach following teacher training. Identification of a successful coaching model could help researchers to support this program at scale, help teachers advance their PD, and improve the outcomes of the children they teach. The results from this cluster RCT suggest that providing teachers with initial training in SAFMEDS and then three 1-hour visits and email contact with a researcher has a positive effect on children's fluency of arithmetic facts compared to initial training only. This paper also contributes to the growing literature reporting the effects of coaching teachers to implement evidence-based interventions within their schools, with a specific focus on mathematics outcomes.

The Education Endowment Foundation (EEF, 2019) are a leading UK charity that support the generation of research and good practice within schools. They aim to support teachers to use evidence that works to improve educational outcomes for children. In their recent implementation guidance report, the EEF identified the importance of reinforcing initial training for interventions with expert follow-on support within school. The results from our study further support this guidance in the context of a SAFMEDS mathematics intervention. Whilst children attending schools in the no support arm of this trial did improve their fluency of arithmetic facts, children made

more significant progress when their teacher received coaching to support their implementation of the SAFMEDS strategy.

Data from Kraft et al. (2018) supported the idea that someone with expertise can coach teachers to implement evidence-based interventions in schools. The results from their meta-analysis revealed that PD programs with an element of coaching can have a positive effect on student achievement outcomes by $+0.18 SD$; although this largely reflected their application to literacy and content-based interventions. In contrast, Lynch et al.'s (2019) meta-analysis suggested that there was no added benefit to having a coaching element as a part of the format for PD interventions for mathematics and science. In the current study, we provided direct experimental manipulation of a coaching element to the SAFMEDS intervention and found the effect of coaching to be between $+0.21$ and $0.23 SD$ across the MFaCTs measures; these outcomes are similar to Kraft et al.'s findings. Our results provide some additional support for the effectiveness of teacher coaching in the context of a fluency-based arithmetic intervention.

Kraft et al.'s (2018) analysis also revealed that effect sizes varied significantly depending on whether the researchers devised their own assessments or administered standardized tests. When considering effect sizes within education research, Kraft (2020) outlined that researcher-designed assessments often reflect content that more closely align with the outcomes of the evaluated program, compared to the broader scope of standardized assessments. The MFaCTs measures are published and standardized, however the focus on fluency of arithmetic facts aligned closely with the content the children covered within the SAFMEDS sessions. This may have inflated the observed effect sizes compared to alternative standardized assessments.

When interpreting the results from the current study, it is important to consider the underlying mechanisms and social contingencies that might have made coaching

effective. First, the support visits served to provide teachers with feedback to improve their implementation fidelity. Durlak and DuPre (2008) reported that without support teachers are often unable to implement an intervention to 100% fidelity following training. This is not surprising given that field studies come with additional extraneous variables compared to efficacy/laboratory designs (Cook & Odom, 2013). However, improved implementation fidelity of evidence-based interventions in the classroom can lead to improved student outcomes (Durlak & DuPre, 2008; Ysseldyke et al., 2003). By providing teachers with in-situ support during the present RCT, our aim was to help them deliver the program in a way that more closely aligns with its intended design (e.g., ensuring the children engaged with the practice regularly, discussing methods to address and reduce cheating, as well as reviewing and acting upon children's progress data). Improved adherence to the procedural aspects of the program may explain why the children who attended schools allocated to the support trial arm made greater fluency gains.

Second, between each visit, the teachers had the opportunity to adapt their practice based on the feedback they had received. By design, PT practices allow teachers to monitor and reflect upon their children's learning (Lindsley, 1995). If children are not making desirable progress towards fluency, then their teacher *should* adapt the instruction or materials that they provide. Through session observation and review of these data, the first author would have been able to see progress across the program. As such, there is a level of accountability that the teachers might have experienced to avoid feeling embarrassed during the following support visit. In a qualitative evaluation of a coaching program in a healthcare setting, Liddy, Johnston, Irving, Nash, and Ward (2015) reported that coaches helped patients realize that they need to play an active role in managing and improving their health. The patients also

reported that their personal accountability increased as a result of their engagement with the coaches because they knew someone else was monitoring their engagement with the program. It seems feasible that this finding could extend to programs relating to school-based educational interventions. It is difficult to disentangle implementation fidelity and accountability, but both of these mechanisms provide direction for future research in this area.

We acknowledge that this study would have been enhanced if we collected data relating to teachers' and children's implementation fidelity across both trial arms. By employing blind observers to attend a SAFMEDS session in each school following each cycle of support visits, it would have been possible to directly assess the effects of expertise on the teacher's implementation fidelity. Moreover, analysis of these data could identify common aspects of the strategy that teachers struggle, or fail, to implement in school settings. Due to practical constraints and funding, we were unable to incorporate this into the current study. However, this would provide a valuable extension to future replications. It is possible that the implementation support offered by the researcher helped teachers to: interpret learning pictures more readily and accurately; identify and correct children's procedural steps as they progressed through each SAFMEDS timing; as well as manage challenges such as cheating and identifying appropriate materials.

Whilst we did not carry out a formal economic analysis, we believe that this support model may be a cost-effective and feasible alternative to embedding a researcher in each school to run and maintain a SAFMEDS intervention. Adoption of the current support model would enable a researcher to provide necessary implementation support at scale and may encourage teachers to use the intervention beyond the termination of a research study. Costs associated with the replication of this

support model include a researcher's time (three 1-hour support visits and designated time to respond to email queries), cost of travel between schools, the cost of materials (e.g., printing each SAFMEDS deck double-sided onto card; at approximately 6 sheets of A4 card per deck per child), and the cost a teacher/TAs time to prepare and deliver three SAFMEDS sessions per week (with each session lasting approximately 20-minutes).

The results from the current study suggest that initial training can provide teachers with skills to implement the SAFMEDS strategy in their school. Children across both trial arms evidenced improvements in their fluency of arithmetic facts between baseline and follow-up across both MFaCTs measures. Support from a researcher helped teachers to elicit greater fluency progress from the children that they worked with. Further research is still needed to establish the components of this model that make the support effective; including the exploration of the effects of coaching on teacher's implementation fidelity and perceived accountability.

Acknowledgements

We would like to thank all the schools, teachers, and children who participated, the Economic and Social Research Council Doctoral Training Partnership for funding the PhD project, the blind assessors who administered the assessments, and the Regional School Effectiveness and Improvement Service for North Wales (GwE) for their support throughout.

Declaration of interest

K.O., S.H., R.W., R.H., and C.H. designed and managed the study. K.O. was responsible for trial management. K.O., J.P., and T.B. contributed to data analysis. K.O. and S.H. devised and delivered the SAFMEDS training. They also both receive occasional fees to deliver SAFMEDS training to schools in the North Wales region and other parts of the UK. R.W. works for the Regional School Effectiveness and Improvement Service for North Wales (GwE), who promoted this project and organizes regular SAFMEDS training for schools across the region. All authors made a contribution to the interpretation of the study findings and/or writing and editing the paper.

Funding

This work was supported by the Economic Social Research Council Doctoral Training Partnership (ESRC-DTP) under the grant number ES/P00069X/1.

Data availability

The data that support the finding of this study are available from the corresponding author, K.O., upon reasonable request.

Word count: 6636

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412. doi: 10.1016/j.jml.2007.12.005
- Beverley, M., Hughes, J. C., & Hastings, R. P. (2018). Improving essential numeracy skills in primary school children using a brief fluency-building intervention: A randomized control trial. *Wales Journal of Education*, 20(1), 114-134. doi: 10.16922/wje.20.1.7
- Beverley, M., Hughes, J. C., & Hastings, R. P. (2016). Using SAFMEDS to assist language learners to acquire second-language vocabulary. *European Journal of Behavior Analysis*, 17(2), 131-141. doi: 10.1080/15021149.2016.1247577
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, 19(2), 163-197. doi: 10.1007/BF03393163
- Binder, C., Haughton, E., & Bateman, B. (2002). Fluency: Achieving true mastery in the learning process. Retrieved from https://www.researchgate.net/profile/Ogholgol_Nazari/publication/269793677_On_Situating_the_Stance_of_Practice_in_SLA/links/56b44f9808ae61c480581e9d.pdf [Accessed 26 June 2020]
- Casey, J., McLaughlin, T. F., Weber, K. P., & Everson, M. (2003). The effects of five-minute practice, unlimited practice, with SAFMED cards on correct and error rate in math facts for two elementary school children with learning disabilities. *International Journal of Special Education*, 18(1), 66-72. doi: 10.1.1.627.1342
- Cihon, T. M., Sturtz, A. M., & Eshleman, J. (2012). The effects of instructor-provided or student-created flashcards with weekly, one-minute timings on unit quiz scores in introduction to applied behavior analysis courses. *European Journal of Behavior Analysis*, 13, 47-57. doi: 10.1080/10521149.2012.11434404

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Ed.). New York: Erlbaum.
- Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children*, 79(2), 135-144. doi: 10.1177/001440291307900201
- Cunningham, D., McLaughlin, T. F., & Weber, K. P. (2012). The use and evaluation of verbal prompting with see to say problems and answers and SAFMEDS to teach math facts to a student with learning disabilities. *International Journal of Advances in Psychology*, 1(2), 37-39.
- Durlak, J. A. & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on programme outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327-350. doi: 10.1007/s10464-008-9165-0
- Education Endowment Foundation (2019). Putting evidence to work: A school's guide to implementation guidance report. *Retrieved from:* https://educationendowmentfoundation.org.uk/public/files/Publications/Implementation/EEF_Implementation_Guidance_Report_2019.pdf [Accessed 13 February 2020]
- Feingold, A. (2015). Confidence interval estimation for standardized effect sizes in multilevel and latent growth modelling. *Journal of Consulting and Clinical Psychology*, 83I(1), 157-168. doi: 10.1037/a0037721
- Fletcher, S., & Mullen, C. A. (Eds.). (2012). *SAGE handbook of mentoring and coaching in education*. Los Angeles, California: Sage.

- Fuller, J. L., & Fienup, D. M. (2018). A preliminary analysis of mastery criterion level: effects on response maintenance. *Behaviour Analysis in Practice*, 11(1), 1-8. doi: 10.1007/s40617-017-0201-0
- Furlong, J. (2015). Teaching tomorrow's teachers. *Read online:*
<https://beta.gov.wales/sites/default/files/publications/2018-03/teaching-tomorrow's-teachers.pdf>
- Galecki, A., & Burzykowski, T. (2013). Linear mixed-effects model. In *Linear Mixed-Effects Models Using R* (pp. 245-273). Springer, New York, NY. doi: 10.1007/978-1-4614-3900-4
- Greene, I., Mc Tiernan, A., & Holloway, J. (2018). Cross-Age Peer Tutoring and Fluency-Based Instruction to Achieve Fluency with Mathematics Computation Skills: A Randomized Controlled Trial. *Journal of Behavioral Education*, 27(2), 145-171. doi: 10.1007/s10864-018-9291-1
- Hunter, S. H., Beverley, M., Parkinson, J., & Hughes, J. C. (2016). Increasing high school students' maths skills with the use of SAFMEDS class-wide. *European Journal of Behavior Analysis*, 17(2), 154-165. doi: 10.1080/15021149.2016.1247643
- Jahoda, A., Hastings, R., Hatton, C., Cooper, S. A., Dagnan, D., Zhang, R., ... & Scott, K. (2017). Comparison of behavioural activation with guided self-help for treatment of depression in adults with intellectual disabilities: a randomised controlled trial. *The Lancet Psychiatry*, 4(12), 909-919. doi: 10.1016/S2215-0366(17)30426-1
- Johnson, K., & Street, E. M. (2012). *Response to intervention and precision teaching: creating synergy in the classroom*. New York: Guilford Press.

- Klingner, J. K., Boardman, A. G., & McMaster, K. L. (2013). What does it take to scale up and sustain evidence-based practices? *Exceptional Children*, 79(2), 195-211. doi: 10.1177/001440291307900205
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241-253. doi: 10.3102/0013189X20912798
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547-588. doi: 10.3102/0034654318759268
- Kubina, R. M., Ward, M. C., & Mozzoni, M. P. (2000). Helping one person at a time: precision teaching and traumatic brain injury rehabilitation. *Behavioral Interventions*, 15(3), 189-203. doi: 10.1002/1099-078X(200007/09)15:3<189::AID-BIN55>3.0.CO;2-X
- Kubina, R. M., & Yurich, K. K. L. (2012). *The Precision Teaching Book*. Lemont, PA: Greatness Achieved.
- Lambe, D., Murphy, C., & Kelly, M. E. (2015). The impact of a precision teaching intervention on the reading fluency of typically developing children. *Behavioral interventions*, 30(4), 364-377. doi: 10.1002/bin.1418
- Lee, O., Penfield, R., & Maerten-Rivera, J. (2009). Effects of fidelity of implementation on science achievement gains among English language learners. *Journal of Research in Science Teaching*, 46(7), 836–859. doi: 10.1002/tea.20335
- Liddy, C., Johnston, S., Irving, H., Nash, K., & Ward, N. (2015). Improving awareness, accountability, and access through health coaching: qualitative study of patients' perspectives. *Canadian Family Physician*, 61(3), 158-164.
- Lindsley, O. R. (1995). Precision Teaching: By teachers for children. *Journal of Precision Teaching*, 12(2), 9-17. doi: 10.1177/004005999002200302

- Lynch, K., Hill, H.C., Gonzalez, K., & Pollard, C. (2019). Strengthening the Research Base that Informs STEM Instructional Improvement Efforts: A Meta-Analysis. *Educational Evaluation and Policy*, 41(3), 260-293. doi: 10.3102/0162373719849044
- OECD (2010). PISA 2009 results: What students know and can do – student performance in reading, mathematics, and science (volume I). doi: 10.1987/9789264091450-en
- OECD (2012), Equity and Quality in Education: Supporting Disadvantaged Students and Schools, OECD Publishing, Paris. doi: 10.1787/9789264130852-en.
- OECD (2016). PISA 2015 results (Volume I): Excellence and equity in education. Paris: OECD Publishing. doi: 10.1787/9789264266490-en
- Quigley, S. P., Peterson, S. M., Frieder, J. E., & Peck, K. M. (2018). A review of SAFMEDS: Evidence for procedures, outcomes and directions for future research. *Perspectives on Behavior Science*, 1-19. doi: 10.1007/s40614-017-0087-8
- Reynolds, C. R., Voress, J. K., & Kamphaus, R. W. (2015). *Mathematics Fluency and Calculation Tests: Elementary*. Texas: Pro-ed
- Sailors, M., & Shanklin, N. L. (2010). Introduction: Growing evidence to support coaching in literacy and mathematics. *Elementary School Journal*, 111(1), 1-6. doi: 10.1086/653467
- Stockwell, F., & Eshleman, J. (2010). A case study using SAFMEDS to promote fluency with Skinner's verbal behavior terms. *Journal of Precision Teaching and Celeration*, 26, 33-40.

- Tyler, E., Watkins, R. C., Roberts, S. E., Hoerger, M., Hulson-Jones, A., Hughes, J. C., & Hastings, R. P. (2019). Research note: Collaborative Institute for Education Research, Evidence and Impact (CIEREI). *Wales Journal of Education*, 20(1), 135-140. doi: 10.16922/wje.20.1.8
- Van Rinsveld, A., Dricot, L., Guillaume, M., Rossion, B., & Schiltz, C. (2017). Mental arithmetic in the bilingual brain: Language matters. *Neuropsychologia*, 101, 17-29. doi: 10.1016/j.neuropsychologia.2017.05.009
- Ysseldyke, J., Spicuzza, R., Kosciulek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance mathematics achievement in Urban schools. *Journal of Education for Students Placed at Risk*, 8(2), 247-265. doi: 10.1207/S15327671ESPR0802_4
- Zimmermann, C., Swami, N., Krzyzanowska, M., Hannon, B., Leighl, N., Oza, A., ... & Donner, A. (2014). Early palliative care for patients with advanced cancer: a cluster-randomised controlled trial. *The Lancet*, 383(9930), 1721-1730. doi: 10.1016/S0140-6736(13)62416-2

Supplementary material***Percentage of missing data***

Table 1A. Percentage of missing data at baseline and follow-up

Time	MFaCTs Measure	Missing (<i>n</i>)	Total (<i>n</i>)	Percentage missing
Baseline	Grades 1-2	21	575	3.65
	Grades 3-5	21	575	3.65
Follow-up	Grades 1-2	82	575	14.26
	Grades 3-5	86	575	14.96

Comparisons between model outcomes (missing data vs complete case vs imputation data sets)

Table 2A. Summary of the model outcomes based on the dataset containing missing data.

	On-going support		No support		Adjusted cases analysis*			ICC	
	<i>n</i>	Marginal mean of raw scores [95% CI]	<i>n</i>	Marginal mean of raw scores [95% CI]	Adjusted difference between change scores [95% CI]	<i>p</i>	<i>d</i> [95% CI]		
MFaCTs: Grades 1-2									
Baseline	285	12.00 [10.28, 13.73]	269	9.02 [7.28, 10.76]	-	-	-	School	0.36
Follow-up	256	22.59 [20.85, 24.34]	237	16.50 [14.72, 18.29]	2.92 [1.22, 4.61]	.001	0.23 [0.06, 0.40]	Children	0.46
MFaCTs: Grades 3-5									
Baseline	285	8.52 [7.00, 10.05]	269	5.94 [4.40, 7.48]	-	-	-	School	0.37
Follow-up	251	19.14 [17.59, 20.69]	238	13.76 [12.19, 15.34]	2.68 [1.21, 4.14]	<.001	0.25 [0.08, 0.42]	Children	0.42

* Model adjusted for the following fixed effects: county, child's predominant language, gender, free school meals status, school year group.

Interaction effect: trial arm x time. Random effects: school, children.

Table 3A. Summary of The Model Outcomes Based on a Complete Case Analysis

	On-going support		No support		Adjusted cases analysis*			ICC	
	<i>n</i>	Marginal mean of raw scores [95% CI]	<i>n</i>	Marginal mean of raw scores [95% CI]	Adjusted difference between change scores [95% CI]	<i>p</i>	<i>d</i> [95% CI]		
MFaCTs: Grades 1-2									
Baseline	248	11.53 [9.71, 13.35]	229	8.77 [6.90, 10.65]	-	-	-	School	0.37
Follow-up	248	22.40 [20.58, 24.22]	229	16.33 [14.46, 18.21]	3.32 [1.60, 5.04]	<.001	0.26 [0.08, 0.44]	Children	0.46
MFaCTs: Grades 3-5									
Baseline	243	8.22 [6.61, 9.82]	230	5.63 [4.00, 7.27]	-	-	-	School	0.38
Follow-up	243	18.95 [17.34, 20.55]	230	13.60 [11.96, 15.24]	2.76 [1.25, 4.27]	<.001	0.26 [0.08, 0.44]	Children	0.42

* Model adjusted for the following fixed effects: county, child's predominant language, gender, free school meals status, school year group.

Interaction effect: trial arm x time. Random effects: school, children.

Table 4A. Summary of The Model Outcomes Based on 50 Iterations of Imputed Data.

	On-going support		No support		Adjusted cases analysis*			ICC	
	<i>n</i>	Marginal mean of raw scores [95% CI]	<i>n</i>	Marginal mean of raw scores [95% CI]	Adjusted difference between change scores [95% CI]	<i>p</i>	<i>d</i> [95% CI]		
MFaCTs: Grades 1-2									
Baseline	294	12.38 [10.71, 14.05]	281	9.05 [7.38, 10.72]	-	-	-	School	0.32
Follow-up	294	22.62 [20.87, 24.36]	281	16.62 [14.87, 18.37]	2.67 [0.75, 4.58]	.006	0.21 [0.04, 0.37]	Children	0.43
MFaCTs: Grades 3-5									
Baseline	294	8.70 [7.25, 10.15]	281	6.00 [4.54, 7.47]	-	-	-	School	0.33
Follow-up	294	19.18 [17.70, 20.67]	281	13.66 [12.15, 15.18]	2.81 [1.20, 4.42]	.001	0.26 [0.09, 0.42]	Children	0.40

* Model adjusted for the following fixed effects: county, child's predominant language, gender, free school meals status, school year group.

Interaction effect: trial arm x time. Random effects: school, children

Table 5A. Baseline measurements for the children's outcome measures, based on the dataset containing missing data

	No Support			Support		
	<i>n</i>	<i>M (SD)</i>	Min to max	<i>n</i>	<i>M (SD)</i>	Min to max
Measure						
MFaCTs: Grades 1-2	269	9.00 (6.88)	0 to 35	285	12.29 (10.26)	0 to 65
MFaCTs: Grades 3-5	269	5.94 (4.83)	0 to 26	285	8.70 (7.89)	0 to 44

Table 6A. Baseline measurements for the children's outcome measures, based on complete case analysis

	No Support			Support		
	<i>n</i>	<i>M (SD)</i>	Min to max	<i>n</i>	<i>M (SD)</i>	Min to max
Measure						
MFaCTs: Grades 1-2	229	8.87 (6.82)	0 to 35	248	11.24 (11.12)	0 to 60
MFaCTs: Grades 3-5	230	5.71 (4.71)	0 to 26	243	7.88 (6.63)	0 to 35

Table 7A. Baseline measurements for the children's outcome measures, based on imputed data set

	No Support			Support		
	<i>n</i>	<i>M (SD)</i>	Min to max	<i>n</i>	<i>M (SD)</i>	Min to max
Measure						
MFaCTs: Grades 1-2	281	9.06 (6.94)	0 to 36.01	294	12.52 (10.52)	0 to 65.28
MFaCTs: Grades 3-5	281	5.99 (4.89)	0 to 26.76	294	8.85 (8.01)	0 to 44.21

Full output tables for data reported in manuscript

Table 8A. Model outcomes for the MFaCTs: Grades 1-2 measure

Outcome: Score	Model outcomes			
	β [95% CI]	SE	Z	$p> z $
Language	1.68 [-0.88, 4.24]	1.31	1.28	.20
County	-0.13 [-0.80, 0.54]	0.34	-0.38	.71
Gender	0.78 [-0.56, 2.13]	0.69	1.14	.25
eFSM	-0.66 [-2.22, 0.91]	0.80	-0.82	.41
Year group	4.52 [3.63, 5.41]	0.45	9.96	<.001
Trial arm	0.46 [-2.05, 2.96]	1.28	0.36	.72
Time	7.59 [6.38, 8.80]	0.62	12.30	<.001
Interaction: Trial arm x Time	2.92 [1.22, 4.61]	0.86	3.38	.001
_cons	-5.12 [-11.34, 1.10]	3.17	-1.61	.11
Random-effects parameters				
	Estimate [95% CI]	SE		
School	3.72 [2.78, 4.97]	0.55		
Children	5.77 [5.11, 6.52]	0.36		
Residual	6.71 [6.30, 7.16]	0.22		

Table 9A. Model outcomes for the MFaCTS: Grades 3-5 measure

Outcome: Score	Model outcomes			
	β [95% CI]	<i>SE</i>	<i>Z</i>	<i>p</i> > <i>z</i>
Language	0.31 [-1.92, 2.54]	1.14	0.27	.79
County	-0.10 [-0.69, 0.50]	0.30	-0.32	.75
Gender	0.57 [-0.49, 1.64]	0.54	1.06	.29
eFSM	-0.83 [-2.07, 0.41]	0.63	-1.31	.19
Year group	3.49 [2.73, 4.25]	0.39	8.99	<.001
Trial arm	0.63 [-1.58, 2.84]	1.23	0.56	.59
Time	7.87 [6.83, 8.91]	0.53	14.82	<.001
Interaction: Trial arm x Time	2.68 [1.21, 4.14]	0.75	3.59	<.001
_cons	-2.79 [-8.13, 2.55]	2.72	-1.02	.31
Random-effects parameters				
	Estimate [95% CI]	<i>SE</i>		
School	3.44 [2.64, 4.49]	0.46		
Children	4.19 [3.63, 4.84]	0.31		
Residual	5.81 [5.45, 6.19]	0.19		

Moderation analyses***Gender (level 1. Male vs female)***

Table 10A. Outcomes for moderation analysis between gender, trial arm, and time

Interaction	MFaCTs: Grades 1-2		MFaCTs: Grades 3-5	
	Adjusted difference between change scores * [95% CI]	<i>p</i>	Adjusted difference between change scores * [95% CI]	<i>p</i>
Gender x trial arm	1.44 [-1.63, 4.57]	.35	1.20 [-1.31, 3.71]	.35
Gender x time	0.01 [-2.40, 2.42]	.99	0.79 [-1.28, 2.86]	.45
Trial arm x time	2.22 [-0.17, 4.60]	.07	2.33 [0.26, 4.49]	.03
Gender x trial arm x time	1.38 [-1.99, 4.76]	.42	0.69 [-2.23, 3.60]	.64

* Model adjusted for the following fixed effects: county, child's predominant language, free school meals status, school year group. Random effects: school, children

Eligibility for free school meals (eFSM) status (Level 1. Eligible vs not eligible).

Table 11A. Outcomes for moderation analysis between a measure of social deprivation (eFSM), trial arm, and time

Interaction	MFaCTs: Grades 1-2		MFaCTs: Grades 3-5	
	Adjusted difference between change scores* [95% CI]	<i>p</i>	Adjusted difference between change scores* [95% CI]	<i>p</i>
eFSM x trial arm	0.36 [-3.21, 3.93]	0.84	-1.15 [-4.04, 1.74]	.43
eFSM x time	-1.18 [-3.92, 1.55]	0.40	-1.06 [-3.42, 1.30]	.38
Trial arm x time	3.27 [1.32, 5.22]	.001	2.97 [1.29, 4.66]	.001
eFSM x trial arm x time	-1.50 [-5.36, 2.37]	.45	-1.28 [-4.62, 2.07]	.45

* Model adjusted for the following fixed effects: county, child's predominant language, gender, school year group. Random effects: school, children.

Predominant language (Level 1. English vs Welsh).

Table 12A. Outcomes for moderation analysis between language, trial arm, and time

Interaction	MFaCTs: Grades 1-2		MFaCTs: Grades 3-5	
	Adjusted difference between change scores*	<i>p</i>	Adjusted difference between change scores*	<i>p</i>
	[95% CI]		[95% CI]	
Language x trial arm	2.12 [-3.30, 7.54]	.44	1.02 [-3.70, 5.74]	.67
Language x time	-0.36 [-0.39, 2.35]	.79	-0.36 [-2.70, 1.98]	.76
Trial arm x time	1.11 [-2.19, 4.41]	.51	1.88 [-0.95, 4.72]	.19
Language x trial arm x time	2.43 [-1.42, 6.28]	.22	1.08 [-2.34, 4.39]	.52

* Model adjusted for the following fixed effects: county, gender, free school meals status, school year group. Random effects: school, children.

Tables

Table 1. Characteristics of schools included in the randomization

	Variable	Ongoing support	No support
		<i>n</i>	<i>n</i>
Stage of school	Primary	31	31
	Secondary (high)	2	-
School type	Mainstream	32	30
	Special educational needs	1	1
County/local authority	Conwy	12	14
	Denbighshire	7	3
	Flintshire	5	4
	Gwynedd	5	4
	Anglesey	1	2
	Wrexham	3	4
Predominant language	English	21	21
	Welsh	12	10

Table 2. Role of teaching staff who received training at the beginning of the project

Job title	Ongoing support (<i>n</i>)	No support (<i>n</i>)
Head teacher	2	1
Deputy head teacher	3	3
Teacher	22	18
HLTA*	1	5
Teaching assistant	30	36
Intervention coordinator	1	-
ALNCo**	-	1

*HLTA = higher level teaching assistant **ALNCo = additional learning needs coordinator

Table 3. Characteristics of the children randomized at baseline

	Variable	Ongoing support	No Support
		<i>n</i>	<i>n</i>
Gender	Male	162	158
	Female	152	143
Eligible for free school meals (eFSM)	Yes	75	83
	No	228	212
Predominant language	English	222	219
	Welsh	92	82
County	Conwy	128	145
	Denbighshire	48	30
	Flintshire	39	37
	Gwynedd	58	40
	Anglesey	18	10
	Wrexham	23	39
School year (age)	2 (6-7 years)	93	163
	3 (7-8 years)	192	128
	4 (8-9 years)	4	9
	5 (9-10 years)	3	1
	6 (10-11 years)	2	-
	7 (11-12 years)	11	-
	8 (12-13 years)	1	-
	9 (13-14 years)	4	-
	10 (14-15 years)	-	-
	11 (15-16 years)	4	-

Table 4. Baseline measurements for the outcome measures based on all of the children included in the randomization

Measure	Ongoing support			No support		
	<i>n</i>	<i>M (SD)</i>	Min to max	<i>n</i>	<i>M (SD)</i>	Min to max
MFaCTs: Grades 1-2	304	10.53 (8.88)	0 to 65	288	10.45 (8.88)	0 to 35
MFaCTs: Grades 3-5	304	7.26 (6.61)	0 to 21	288	7.19 (6.60)	0 to 26

Table 5. An outline of the Say-All-Fast-Minute-Every-Day-Shuffled strategy

Timing	Action	Learning principle [corresponding action]
Before timing	1. Shuffle the cards.	Prevents serial learning [1]
	2. Teacher sets a timer for 1-minute.	Short focused practice sprints [2]
During timing	3. Children read the front of the card in their head and say the answer out loud. They should turn each card over to check their answer, before placing it in either their “correct” or “not yet” pile.	Active responding [3] Immediate feedback [3]
After timing	4. Once the timer has finished, the teacher says stop.	
	5. Children count their cards and write their scores down in the given table.	Formative assessment [5]
	6. If a child gets any cards in their “not yet” pile, they should address these cards (error correction).	Practice and firm new skills [6]
	7. All cards should be put back in one pile, ready to shuffle and go again.	
	8. Following all four timings, the child should take their best score and plot it on their SCC.	Repetition to build mastery [7]
	9. At the end of each week, a teacher should look at each child’s data. If they have shown little, to no, progression over three consecutive days they should consider making a change within the program.	Assessment of learning [8,9]

Table 6. Summary of the model outcomes

	Ongoing support		No support		Adjusted cases analysis*			ICC	
	<i>n</i>	Marginal mean of raw scores [95% CI]	<i>n</i>	Marginal mean of raw scores [95% CI]	Adjusted difference between change scores [95% CI]	<i>p</i>	<i>d</i> [95% CI]		
MFaCTs: Grades 1-2									
Baseline	285	12.00 [10.28, 13.73]	269	9.02 [7.28, 10.76]	-	-	-	School	0.36
Follow-up	256	22.59 [20.85, 24.34]	237	16.50 [14.72, 18.29]	2.92 [1.22, 4.61]	.001	0.23 [0.06, 0.39]	Children	0.46
MFaCTs: Grades 3-5									
Baseline	285	8.52 [7.00, 10.05]	269	5.94 [4.40, 7.48]	-	-	-	School	0.37
Follow-up	251	19.14 [17.59, 20.69]	238	13.76 [12.19, 15.34]	2.68 [1.21, 4.14]	<.001	0.25 [0.08, 0.43]	Children	0.42

* Model adjusted for the following fixed effects: county, child's predominant language, gender, free school meals status, school year group. Interaction effect: trial arm x time. Random effects: school, children.

Figures

Figure 1. CONSORT flow diagram

